NONPARAMETRIC FUNCTION ESTIMATION OF THE RELATIONSHIP BETWEEN TWO REPEATEDLY MEASURED VARIABLES

A. Ruckstuhl, A. H. Welsh and R. J. Carroll *

January 13, 1997

Abstract

We describe methods for estimating the regression function nonparametrically and for estimating the variance components in a simple variance component model which is sometimes used for repeated measures data or data with a simple clustered structure. We consider a number of different ways of estimating the regression function. The main results are that the simple pooled estimator which treats the data as independent performs very well asymptotically but that we can construct estimators which perform better asymptotically in some circumstances.

Key words and phrases: Local linear regression, local quasi-likelihood estimator, smoothing, semiparametric estimation, variance components.

Short title. Nonparametric Regression with Repeated Measures

*Andreas Ruckstuhl is Lecturer and Alan H. Welsh is Reader, Department of Statistics, Australian National University, Canberra ACT 2601. R. J. Carroll is Professor of Statistics, Nutrition and Toxicology, Department of Statistics, Texas A&M University, College Station, TX 77843–3143. Our research was supported by a grant from the National Cancer Institute (CA–57030). Carroll's research was partially completed while visiting the Institut für Statistik und Ökonometrie, Sonderforschungsbereich 373, Humboldt Universität zu Berlin, with partial support from a senior Alexander von Humboldt Foundation research award.

1 INTRODUCTION

In this paper, we consider the semiparametric model

$$Y_{ij} = \alpha_i + m(X_{ij}) + \epsilon_{ij}, i = 1, ..., n, j = 1, ..., = J, \tag{1}$$ {eq:qa1}

where α_i and ϵ_{ij} are independent mean zero random variables with variances $\sigma_\alpha^2 > 0$ and $\sigma_\epsilon^2 > 0$, respectively, and $m(\cdot)$ is an unknown smooth function. Let $\underline{\mathbf{Y}}_i = (Y_{i1}, ..., Y_{iJ})^t$, $\underline{\mathbf{X}}_i = (X_{i1}, ..., X_{iJ})^t$, and $\underline{\mathbf{m}}(\underline{\mathbf{X}}_i) = \{m(X_{i1}), ..., m(X_{iJ})\}^t$. The model implies that the $\underline{\mathbf{Y}}_i$ are independent with $E(\underline{\mathbf{Y}}_i|\underline{\mathbf{X}}_i)$ $= \underline{\mathbf{m}}(\underline{\mathbf{X}}_i)$ and, if $\underline{\mathbf{e}}_J = (1, ..., 1)^t$ is the J-vector of ones, $\text{cov}(\underline{\mathbf{Y}}_i|\underline{\mathbf{X}}_i) = \underline{\Sigma} = \sigma_\alpha^2 \underline{\mathbf{e}}_J \underline{\mathbf{e}}_J^t + \sigma_\epsilon^2 \underline{\mathbf{I}}$. We address the general problem of how, when J is fixed and typically small, to estimate the function $m(\cdot)$ nonparametrically and, at the same time, how to estimate the variances σ_α^2 and σ_ϵ^2.

We will show in Section 5 that the variance components $(\sigma_\alpha^2, \sigma_\epsilon^2)$ can be estimated at the parametric rate $\mathcal{O}_P(n^{-1/2})$ and thus can effectively be treated as known for the purpose of developing and analysing estimators of $m(\cdot)$. We therefore treat both variances as known for our theoretical investigation in Sections 2 - 4. For definiteness, we focus on the use of local linear kernel smoothing. Where local linear kernel smoothing yields surprising results (Section 3), we compare these results with results obtained using kernel (local average kernel smoothing) and local quadratic kernel smoothing.

In Section 2, we investigate two simple approaches to the problem of estimating $m(x_0)$ at a fixed point x_0. The first, which we call *pooled estimation*, ignores the dependence structure in the model (1) and simply fits a single nonparametric regression model (with a bandwidth depending on x_0 but not j) through all the data. The second approach, which we call *component estimation*, involves fitting separate nonparametric regression models relating the jth component of $\underline{\mathbf{Y}}$ to the jth component of $\underline{\mathbf{X}}$ (allowing different local bandwidths at x_0 for each component, $j = 1, ..., J$) and then combining these estimators to produce an overall estimator of the common regression function $m(x_0)$.

Pooled estimation has the advantage of simplicity, since only one regression fit is required. Component estimation requires J regression fits and may be adversely affected by boundary effects: if the support of the components of $\underline{\mathbf{X}}$ depends on j, the components estimators may end up combining estimators from components affected by boundary effects with estimators from components unaffected by them. However, we show here that for local linear kernel estimation, pooled estimation is asymptotically equivalent to the optimal linear combination of the component estimators. The property on which this result depends is that for local polynomial kernel regression, the estimators of the component functions are asymptotically independent. (The well-known correspondence between local polynomial kernel regression with local bandwidths and local polynomial nearest neighbor (loess) regression suggests that the same asymptotic independence results holds for the latter.)

Severini & Staniswalis (1994) introduced quasi-likelihood estimation for so-called *partially linear models*, which consist of a linear parametric component, a nonparametric component, and a general covariance structure. Hence model (1) is a simple special case of a partially linear model. We discuss quasi-likelihood estimation in the context of model (1) in Section 3. Severini and Staniswalis focus their analysis on the problem of estimating and deriving asymptotic results for estimators of the parameters of the parametric component of a partially linear model, while we derive asymptotic results for the estimator of the nonparametric component based on local polynomial estimators restricting, however, our attention to simpler models like (1). Since the calculations yield a complicated expression for the asymptotic variance, preventing direct comparisons of estimation methods, we explore in detail the case of independent and identically distributed explanatory variables X. In this case, the asymptotic variance of the locally linear quasi-likelihood estimator is larger than that of the pooled estimator. We found this result surprising so explored the properties of kernel and local quadratic kernel smoothing in quasi-likelihood estimation. We found that (i) the asymptotic variance of the locally linear quasi-likelihood estimator is even larger than that of a (locally averaged) kernel quasi-likelihood estimator (without the bias necessarily being smaller) and (ii) the asymptotic variance of a locally quadratic quasi-likelihood estimator is of a different order than that of the locally linear quasi-likelihood estimator, namely of order $\mathcal{O}_P\left\{(nh^5)^{-1}\right\}$. The increase in the size of the variance of the locally linear quasi-likelihood estimator compared to that of the pooled estimator is caused by the off-diagonal elements of the inverse covariance matrix. In Section 3 we also show that a modified version of the quasi-likelihood estimator in which the inverse covariance matrix is replaced by the diagonal matrix with the diagonal of the inverse covariance matrix on its diagonal, results in an estimator which is asymptotically equivalent to the pooled estimator.

Although the pooled estimator is the (asymptotically) best estimator we have considered so far and is easy to apply, it makes no use of the covariance structure in the components of $\underline{\mathbf{Y}}$ and therefore ought to be capable of being improved upon. Because of the local nature of nonparametric regression, constructing an estimator which accounts for the covariance structure and improves upon the pooled estimator is a surprisingly difficult task (cf. Section 3). In Section 4 we propose a *two-step estimator*. The intuition for it is very simple: in model (1), multiply both sides of the model by the square-root of the inverse covariance matrix and rearrange terms so that we have *"expression"*= $m(\underline{\mathbf{X}}_{ij}) + \underline{\xi}_{ij}$, where the ξ_{ij} are now independent and identically distributed. The *"expression"* depends on $m(\underline{\mathbf{X}}_{ij})$ which we estimate by the pooled estimator. The two-step estimator has a smaller asymptotic variance than the pooled estimator and an asymptotic bias which can be smaller than the pooled estimator.

We require the following assumptions.

C–1 $K(\cdot)$ is a symmetric, compactly supported, bounded kernel density function with unit variance and define $K_h(v) = h^{-1}K(v/h)$, with bandwidth h. Let $\mu(r) = \int z^r K(z)dz$ and $\gamma(r) = \int z^r K^2(z)dz < \infty, r = 1, 2$, with $\gamma = \gamma(0) > 0$.

C–2 $h \to 0$ as $n \to \infty$ such that $nh \to \infty$.

C–3 $m(\cdot)$ has continuous second derivatives.

C–4 x_0 is an interior point of the support of the distribution of $\underline{\mathbf{X}}_i$, the density of X_{ij} is twice continuously differentiable, and $f_j(\cdot)$, the marginal density of X_{ij}, satisfies $f_j(x_0) > 0$.

For local linear quasi-likelihood estimation we require longer expansions (and hence stronger conditions) than for the other estimation methods. In this case, we replace conditions $C3 - C4$ by the following stronger conditions.

C–5 $m(\cdot)$ has continuous fifth derivatives.

C–6 x_0 is an interior point of the support of the distribution of $\underline{\mathbf{X}}_i$, the density of X_{ij} is continuously differentiable, the marginal density f_j of X_{ij} is twice continuously differentiable and satisfies $f_j(x_0) > 0$, and the bivariate joint density f_{jk} of X_{ij} is twice continuously differentiable. Moreover, we require

$$\begin{aligned}
\int (x_j - x_0)^2 f_{j,k}(x_j,\, x_0)dx_j &< \infty \\
\int (x_j - x_0)^r \frac{\partial^\ell}{(\partial x_0)^\ell} f_{j,k}(x_j, x_0)dx_j &< \infty, \quad l = 1, 2,\ r = 1, 2.
\end{aligned}$$

It is sometimes helpful to frame results in the context of an arbitrary covariance matrix. When this is the case, the covariance matrix of $\underline{\mathbf{Y}}$ given $\underline{\mathbf{X}}$ is denoted $\underline{\Sigma} = (\sigma_{j,k})$, the inverse covariance matrix is denoted $\underline{\mathbf{V}} = \underline{\Sigma}^{-1} = (v_{j,k})$, and we set $v_{k\cdot} = \sum_{j=1}^J v_{k,j}$. Recall that under the variance component model, the covariance matrix of $\underline{\mathbf{Y}}$ given $\underline{\mathbf{X}}$ is $\underline{\Sigma} = \sigma_\epsilon^2 \underline{\mathbf{I}} + \sigma_\alpha^2 \underline{\mathbf{e}}_J \underline{\mathbf{e}}_J^t$ so

$$\begin{aligned}
\underline{\mathbf{V}} &= \underline{\Sigma}^{-1} = \sigma_\epsilon^{-2}\left\{\underline{\mathbf{I}} - (d_J/J)\underline{\mathbf{e}}_J\underline{\mathbf{e}}_J^t\right\}; \\
\underline{\mathbf{V}}^{1/2} &= \sigma_\epsilon^{-1}\left(\underline{\mathbf{I}} - \left[\left\{1 - (1 - d_J)^{1/2}\right\}/J\right]\underline{\mathbf{e}}_J\underline{\mathbf{e}}_J^t\right),
\end{aligned}$$

where $d_J = J\sigma_\alpha^2/(\sigma_\epsilon^2 + J\sigma_\alpha^2)$. Under the variance component model, the diagonal and off-diagonal elements of these matrices are constant so it is convenient to denote the diagonal elements $\sigma_{j,j}$ of $\underline{\Sigma}$ by σ_d^2, the diagonal elements $v_{j,j}$ and off-diagonal elements $v_{j,k}$ of $\underline{\mathbf{V}} = \Sigma^{-1}$ as v_d and v_o respectively, and the diagonal and off-diagonal elements of $V^{1/2}$ by $\tilde{v}_d$ and $\tilde{v}_o$ respectively. Finally, under the variance component model, $v_{k\cdot}$ is also constant so we write $v_{k\cdot} = v_{\cdot\cdot}$

2 POOLED AND COMPONENT ESTIMATION

{sec:component}

The pooled estimator $\widehat{m}_{pool}(x_0, h)$ of $m(x_0)$ is defined as the local linear kernel regression estimator with kernel function $K(\cdot)$ and bandwidth h when all the Y's and X's are combined into a single data set of length nJ. That is

$$\begin{aligned}\widehat{m}_{pool}(x_0) &= (1,0)\left\{n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\begin{pmatrix}1\\(X_{ij}-x_0)/h\end{pmatrix}\begin{pmatrix}1\\(X_{ij}-x_0)/h\end{pmatrix}^t K_h(X_{ij}-x_0)\right\}^{-1}\\&\quad\times\left\{n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\begin{pmatrix}1\\(X_{ij}-x_0)/h\end{pmatrix}Y_{ij}K_h(X_{ij}-x_0)\right\}.\end{aligned}$$

The optimal pooled estimator minimizes the mean squared error of $\widehat{m}_{pool}(x_0, h)$ at x_0 over h.

To define the components estimator $\widehat{m}_W(x_0, \underline{\mathbf{h}}, \underline{\mathbf{c}})$, for $j = 1, ..., J$, let $\widehat{m}_j(x_0, h)$ be the local linear kernel regression estimator of the (Y_{ij}) on the (X_{ij}), with bandwidth h_j. That is

$$\begin{aligned}\widehat{m}_j(x_0, h_j) &= (1,0)\left\{n^{-1}\sum_{i=1}^{n}\begin{pmatrix}1\\(X_{ij}-x_0)/h_j\end{pmatrix}\begin{pmatrix}1\\(X_{ij}-x_0)/h_j\end{pmatrix}^t K_{h_j}(X_{ij}-x_0)\right\}^{-1}\\&\quad\times\left\{n^{-1}\sum_{i=1}^{n}\begin{pmatrix}1\\(X_{ij}-x_0)/h_j\end{pmatrix}Y_{ij}K_{h_j}(X_{ij}-x_0)\right\}.\end{aligned}$$

Then if $\underline{\mathbf{h}} = (h_1, ..., h_J)^t$ and $\underline{\mathbf{c}} = (c_1, ..., c_J)^t$, the components estimator is the weighted average of the component estimators given by

$$\widehat{m}_W(x_0, \underline{\mathbf{h}}, \underline{\mathbf{c}}) = \sum_{j=1}^{J} c_j \widehat{m}_j(x_0, h_j), \quad \sum_{j=1}^{J} c_j = 1. \tag{2}$$ {eq:qc1}

The optimal components estimator minimizes the mean squared error at x_0 over both $\underline{\mathbf{h}}$ and $\underline{\mathbf{c}}$.

The following result is proved in appendix A.1 and A.2.

Theorem 1 *Suppose that conditions $C1 - C4$ hold. Define $s(x_0) = J^{-1}\sum_{j=1}^{J} f_j(x_0)$. For local linear kernel regression, the optimal pooled estimator and the optimal components estimator are asymptotically equivalent. The bias, variance, optimal bandwidth and mean squared error at this* *optimal bandwidth for the former are given by*

$$bias\{\widehat{m}_{pool}(x_0,h)\} \approx (1/2)h^2 m^{(2)}(x_0); \quad (3)$$ {eq:qfg1}

$$var\{\widehat{m}_{pool}(x_0,h)\} \approx \gamma(\sigma_\alpha^2+\sigma_\epsilon^2)\{nhJs(x_0)\}^{-1}; \quad (4)$$ {eq:qfg2}

$$h^5_{opt,pool}(local\ linear) \approx \frac{\gamma(\sigma_\alpha^2+\sigma_\epsilon^2)}{nJs(x_0)\{m^{(2)}(x_0)\}^2}; \quad (5)$$ {eq:qfg3}

$$mse_{opt,pool}(local\ linear) \approx (5/4)\{m^{(2)}(x_0)\}^{2/5}\left[\gamma(\sigma_\alpha^2+\sigma_\epsilon^2)\left\{nJs(x_0)\right\}^{-1}\right]^{4/5}. \quad (6)$$ {eq:qfg4}

{th:pooled}

Theorem 1 shows that the pooled estimator has the same asymptotic properties under the model (1) as it has under the nonparametric regression model in which the errors are independent and identically distributed with variance $\sigma_\alpha^2+\sigma_\epsilon^2$. Moreover, working componentwise as in the components estimator does not enable us to make use of the known dependence structure in the model (1) in the sense that we can do no better than using the pooled estimator.

3 QUASI-LIKELIHOOD ESTIMATION

{sec:qle}

In this section we apply Severini & Staniswalis' (1994) proposal to model (1). Because of its undesirable asymptotic properties we modify the quasi-likelihood estimator in the second part of this section. The modification yields an estimator asymptotically equivalent to the pooled estimator.

3.1 Ordinary Quasi-likelihood Estimator

{subsec:sev-stan}

Recall that $\underline{\mathbf{V}} = \underline{\Sigma}^{-1}$. Then $\widehat{m}_{p,qle}(x_0,h)$, the intercept in the solution of

$$\sum_{i=1}^{n}\begin{bmatrix} 1 & \dots & 1 \\ . & \dots & . \\ . & \dots & . \\ (X_{i1}-x_0)^p & \dots & (X_{iJ}-x_0)^p \end{bmatrix} \underline{\mathbf{V}} \begin{bmatrix} K_h(X_{i1}-x_0)\{Y_{i1}-\sum_{k=0}^{p}\widehat{\beta}_k(X_{i1}-x_0)^k\} \\ . \\ . \\ K_h(X_{iJ}-x_0)\{Y_{iJ}-\sum_{k=0}^{p}\widehat{\beta}_k(X_{iJ}-x_0)^k\} \end{bmatrix} = \underline{\mathbf{0}}, \quad (7)$$ {eq:qle}

is the local polynomial version of the quasi-likelihood estimator in Severini & Staniswalis (1994, Equation (18)) for model (1). The local linear quasi-likelihood estimator which we consider first has $p=1$.

In Appendix A.3, we prove the following asymptotic results:

Theorem 2 *Suppose that conditions $C1-C2$ and $C5-C6$ hold. Let $\widehat{m}_{p,qle}(x_0,h)$ be the solution of (7). Then*

$$bias\{\widehat{m}_{1,qle}(x_0,h)\} \approx h^2 m^{(2)}(x_0)/2,$$

$$var\{\widehat{m}_{1,qle}(x_0,h)\} \;=\; \mathcal{O}_P\left\{(nh)^{-1})\right\}.$$

For the variance component model (1), where the X_k's are independent and identically distributed with marginal distribution $f(\cdot)$ and variance σ_X^2, the asymptotic variance reduces to

$$var\{\widehat{m}_{1,qle}(x_0,h)\} \;\approx\; \frac{\gamma(0)\,(\sigma_\alpha^2+\sigma_\epsilon^2)}{nhJf(x_0)}\left\{1+\left(\frac{\eta}{J-\eta}\frac{f^{(1)}(x_0)}{f(x_0)}\right)^2 (J-1)\sigma_X^2\right\}.$$

where $\eta = J\sigma_\alpha^2/(J\sigma_\alpha^2+\sigma_\epsilon^2)$. {th:qle}

Note, that the local linear quasi-likelihood estimator has the same asymptotic bias but a larger asymptotic variance than that of the pooled estimator in the variance component model with independent and identically distributed X_j's. In view of this result (and the lack of results on estimating the nonparametric component of the model referred to in the introduction), we also obtained results for kernel and local quadratic quasi-likelihood estimators ($p=0$ and $p=2$ respectively) of the regression function.

Under similar conditions to those in Theorem 2, we show in Appendix A.3 that the kernel estimator has

$$\begin{aligned}
\text{bias}\{\widehat{m}_{0,qle}(x_0,h)\} \;&\approx\; h^2\left\{m^{(1)}(x_0)\frac{\sum_{k=1}^J v_{k\cdot} f_k^{(1)}(x_0)}{\sum_{k=1}^J v_{k\cdot} f_k(x_0)} + m^{(2)}(x_0)/2\right\},\\
\text{var}\{\widehat{m}_{0,qle}(x_0,h)\} \;&\approx\; \frac{\gamma(0)\left\{\sum_{k=1}^J (v_{k\cdot})^2\sigma_{k,k} f_k(x_0)\right\}}{nh\left(\sum_{k=1}^J v_{k\cdot} f_k(x_0)\right)^2},
\end{aligned}$$

where $v_{k\cdot} = \sum_{j=1}^J v_{k,j}$ with $v_{j,k}$ the elements of $\underline{\mathbf{V}}$. For a variance component model (1), where the X_k's are independent and identically distributed with marginal distribution $f(\cdot)$ and variance σ_X^2, the asymptotic variances reduce to

$$\text{var}\{\widehat{m}_{0,qle}(x_0,h)\} \;\approx\; \frac{\gamma(0)(\sigma_\alpha^2+\sigma_\epsilon^2)}{nh\,J\,f(x_0)}.$$

The kernel quasi-likelihood estimator has the same asymptotic variance as the pooled estimator. However, as it is generally dependent on the design, its bias also depends on the design.

For the local quadratic quasi-likelihood estimator, we show that

$$\text{bias}\{\widehat{m}_{2,qle}(x_0,h)\} \;\approx\; h^4\frac{\mu(6)-\mu^2(4)}{\mu(4)-1}\left\{\frac{m^{(3)}(x_0)}{3!}\frac{S_N}{S_D} - \frac{m^{(4)}(x_0)}{4!}\right\},$$

where S_D and S_N are given in (23) and (24), respectively, and for the variance component model (1), where the X_k's are independent and identically distributed with marginal distribution $f(\cdot)$ and variance σ_X^2, the asymptotic variance is

$$\text{var}\{\widehat{m}_{2,qle}(x_0,h)\} \;\approx\; \frac{\gamma(0)\,(\sigma_\alpha^2+\sigma_\epsilon^2)}{nh^5Jf(x_0)}\left(\frac{\eta}{J-\eta}\right)^2\frac{(J-1)}{\{\mu(4)-1\}^2}\left\{\int (x-x_0)^4 f(x)\,dx - \left(\sigma_X^2\right)^2\right\}.$$

where $\eta = J\sigma_\alpha^2/(J\sigma_\alpha^2 + \sigma_\epsilon^2)$. The asymptotic variance of the local quadratic estimator is of higher order $\mathcal{O}_P(n^{-1}h^{-5})$ than $\mathcal{O}_P(n^{-1}h^{-1})$ obtained by the pooled estimator.

3.2 Modified Quasi-likelihood Estimator

{subsec:mqle}

Analysing the proof of the asymptotic results for the quasi-likelihood estimator, the slow rate of convergence of the asymptotic variance is caused by the off-diagonal elements of $\underline{\mathbf{V}} = \underline{\Sigma}^{-1}$. This suggests that we modify the quasi-likelihood estimator by replacing $\underline{\mathbf{V}} = \underline{\Sigma}^{-1}$ by $\underline{\mathbf{V}} = \mathrm{diag}(\underline{\Sigma}^{-1})$.

Theorem 3 *Suppose the conditions of Theorem 2 hold. Let $\widehat{m}_{1,mqle}(x_0,h)$ be the solution of (7) for $p=1$, where $\underline{\mathbf{V}} = \underline{\Sigma}^{-1} = (v_{j,k})$ is replaced by $\underline{\mathbf{V}} = diag(\underline{\Sigma}^{-1})$. Then*

$$\begin{aligned} bias\{\widehat{m}_{1,mqle}(x_0,h)\} &\approx h^2 m^{(2)}(x_0)/2, \\ var\{\widehat{m}_{1,mqle}(x_0,h)\} &\approx \frac{1}{nh}\frac{\sum_{j=1}^J v_{j,j}^2 \sigma_{j,j} f_j(x_0)}{\left\{\sum_{j=1}^J v_{j,j} f_j(x_0)\right\}^2}\,\gamma(0). \end{aligned}$$

where $\underline{\Sigma} = (\sigma_{j,k})$. For the variance component model (1), the asymptotic variances reduce to

$$var\{\widehat{m}_{1,mqle}(x_0,h)\} \approx \frac{1}{nh}\frac{(\sigma_\alpha^2 + \sigma_\epsilon^2)}{\sum_{j=1}^J f_j(x_0)}\,\gamma(0)$$

and hence the local linear modified quasi-likelihood estimator is asymptotically equivalent to the pooled estimator.

{th:mqle}

For the kernel and local quadratic modified quasi-likelihood estimators, we obtain

$$\begin{aligned} \mathrm{bias}\{\widehat{m}_{0,mqle}(x_0,h)\} &\approx h^2\left\{m^{(1)}(x_0)\frac{\sum_{j=1}^J v_{j,j} f_j^{(1)}(x_0)}{\sum_{j=1}^J v_{j,j} f_j(x_0)} + m^{(2)}(x_0)/2\right\}, \\ \mathrm{bias}\{\widehat{m}_{2,mqle}(x_0,h)\} &\approx h^4\frac{\mu^2(4)-\mu(6)}{\mu(4)-1}\left\{\frac{m^{(3)}(x_0)}{3!}\frac{\sum_{j=1}^J v_{j,j} f_j^{(1)}(x_0)}{\sum_{j=1}^J v_{j,j} f_j(x_0)} + \frac{m^{(4)}(x_0)}{4!}\right\}, \end{aligned}$$

and

$$\mathrm{var}\{\widehat{m}_{p,mqle}(x_0,h)\} \approx \frac{1}{nh}\frac{\sum_{j=1}^J v_{j,j}^2 \sigma_{j,j} f_j(x_0)}{\left\{\sum_{j=1}^J v_{j,j} f_j(x_0)\right\}^2}\,F_p,$$

where $\underline{\Sigma} = (\sigma_{j,k})$ and

$$\begin{aligned} F_p &= \gamma(0) \text{ for } p=0, \\ F_p &= \gamma(0)\frac{2\mu(4)\{\gamma(0)-\gamma(2)\} + \gamma(4) - \gamma(0)}{\{\mu(4)-1\}^2} \text{ for } p=2. \end{aligned}$$

For the variance component model (1), the asymptotic variances reduce to

$$\mathrm{var}\{\widehat{m}_{p,mqle}(x_0,h)\} \;\approx\; \frac{1}{nh}\frac{(\sigma_\alpha^2+\sigma_\epsilon^2)}{\sum_{j=1}^J f_j(x_0)}\,F_p.$$

For $p=2$, the modified quasi-likelihood estimator is asymptotically better than the quasi-likelihood estimator because its variance converges at a faster rate. For $p=0$, it is easy to see that a sufficient condition for asymptotic equivalence is $v_{j\cdot}/v_{j,j} = constant$ which is, for example, satisfied by the variance component model (1).

Note that both the asymptotic bias and the asymptotic variance are invariant to multiplying V by a constant; i.e., the matrix V has to be determined only up to a multiplicative factor.

4 TWO-STEP ESTIMATION

{sec:combine}

In this section, we propose a two-step estimator which exhibits some asymptotic improvement over the pooled and modified quasi-likelihood estimators.

Again let $\underline{\mathbf{V}} = \underline{\Sigma}^{-1}$ and let $\underline{\mathbf{V}}^{1/2}$ be its symmetric square root. Let $\underline{\mathbf{L}} = \tau\underline{\mathbf{V}}^{1/2}$ and let $\widehat{m}_{1,pool}(\cdot)$ be the pooled estimator of Section 2. Write

$$\underline{\mathbf{Z}} = \underline{\mathbf{LY}} - (\underline{\mathbf{L}} - \underline{\mathbf{I}})\underline{\widehat{\mathbf{m}}}_{1,pool}(\underline{\mathbf{X}}).$$

We propose to estimate $m(x_0)$ by $\widehat{m}_C(x_0)$, the local linear kernel regression estimator of the regression of the Z's on the X's, That is, by solving

$$\begin{aligned}\widehat{m}_C(x_0) &= (1,0)\left\{n^{-1}\sum_{i=1}^n\sum_{j=1}^J \begin{pmatrix}1\\(X_{ij}-x_0)/h\end{pmatrix}\begin{pmatrix}1\\(X_{ij}-x_0)/h\end{pmatrix}^t K_h(X_{ij}-x_0)\right\}^{-1}\\ &\quad\times\left\{n^{-1}\sum_{i=1}^n\sum_{j=1}^J \begin{pmatrix}1\\(X_{ij}-x_0)/h\end{pmatrix} Z_{ij}K_h(X_{ij}-x_0)\right\}.\end{aligned}$$

The intuition for this estimator is very simple: write $Y = m(X)+\epsilon$, multiply both sides by $\Sigma^{-1/2}$ and rearrange terms so that we have "*expression*" $= \underline{\mathbf{m}}(\underline{\mathbf{X}}) + \underline{\xi}$, where the ξ_i are now independent and identically distributed. The "*expression*" we obtain equals Z.

In the appendix, we prove the following result:

Theorem 4 *Suppose the conditions of Theorem 1 hold. Define $d_J = J\sigma_\alpha^2/(\sigma_\epsilon^2 + J\sigma_\alpha^2)$. Then, for $\tau > 0$,*

$$bias\{\widehat{m}_c(x_0,h)\} \;\approx\; -1/2h^2 m^{(2)}(x_0)\tau\tilde{v}_d$$

$$-(1/2)h^2\tau\tilde{v}_o \frac{\sum_{j=1}^J f_j(x_0)\sum_{k\neq j} E\{m^{(2)}(X_{1k})|X_{1j}=x_0\}}{\sum_{j=1}^J f_j(x_0)}; \qquad (8)$$ {eq:qd1}

$$var\{\widehat{m}_c(x_0,h)\} \approx (nh)^{-1}\gamma\tau^2 \left\{\sum_{j=1}^J f_j(x_0)\right\}^{-1}, \qquad (9)$$ {eq:qd2}

where $\tilde{v}_d$ and $\tilde{v}_o$ are the diagonal and the off-diagonal elements of $V^{1/2}$, respectively. {th:comb}

An optimal τ can be determined by minimizing the asymptotic mean squared error. The minimization problem results in a cubic equation in τ. Note that since the bias is design dependent (because of the structure of Z), so also is the optimal τ.

If we choose τ equal to σ_ϵ^{-1} then the asymptotic variance of the two-step estimator is smaller than that of the pooled estimator. If we wish to treat $\underline{\mathbf{Z}} = \underline{\mathbf{m}}(\underline{\mathbf{X}}) + \underline{\xi}$ in the same scale as the original data (1), we should set τ equal to $\tilde{v}_d^{-1}$. In this case,

$$\tau^2 = \sigma_\epsilon^2\left[1-\left\{1-(1-d_J)^{1/2}\right\}/J\right]^{-2},$$

where $d_J = J\sigma_\alpha^2/(\sigma_\epsilon^2 + J\sigma_\alpha^2)$. Thus the asymptotic variance is

$$\mathrm{var}\{\widehat{m}_c(x_0,h)\} \approx \frac{(nh)^{-1}\gamma\sigma_\epsilon^2}{\sum_{j=1}^J f_j(x_0)\left[1-\{1-(1-d_J)^{1/2}\}/J\right]^2}.$$

Now

$$\sigma_\epsilon^2 \le \sigma_\epsilon^2[1-\{1-(1-d_J)^{1/2}\}/J]^{-2} \le \sigma_\epsilon^2 + \sigma_\alpha^2$$

so the two-step estimator with $\tau = \tilde{v}_o^{-1}$ has larger asymptotic variance than the two-step estimator with $\tau = \sigma_\epsilon^{-1}$ but still has smaller asymptotic variance than the pooled estimator. In either case, the asymptotic biases of the two estimators are difficult to compare but note that the asymptotic bias of the two-step estimator can be smaller than that of the pooled estimator because $\tilde{v}_o$ is negative, allowing the possibility of cancellation to occur.

5 ESTIMATION OF THE VARIANCE COMPONENTS

{sec:varcomp}

Let $\underline{\mathbf{Y}} = (\underline{\mathbf{Y}}_1^t, ..., \underline{\mathbf{Y}}_n^t)^t$ be the vector of pooled responses and let $\underline{\mathbf{E}}$ be the deviations of $\underline{\mathbf{Y}}$ from the regression line $\{\underline{\mathbf{m}}^t(\underline{\mathbf{X}}_1), ..., \underline{\mathbf{m}}^t(\underline{\mathbf{X}}_n)\}^t$. For all estimators used in this paper, there is an $(nJ)\times(nJ)$ matrix $\underline{\mathbf{S}}$ with the property that the vector of predicted values equals $\underline{\mathbf{SY}}$, and hence the vector of residuals is $(\underline{\mathbf{I}} - \underline{\mathbf{S}})\underline{\mathbf{Y}} = \underline{\mathbf{D}}_*$; explicit formulae in special cases are given in the appendix.

The simplest approach to estimating the variance components is to pretend that the residuals have mean zero and covariance matrix the same as if $m(\cdot)$ were known. For example, the Gaussian "likelihood" for $\tau_\epsilon = \sigma_\epsilon^2$ and $\tau_\alpha = \sigma_\epsilon^2 + J\sigma_\alpha^2$ can be written as

$$-n(J-1)\log\tau_\epsilon - n\log\tau_\alpha - \frac{1}{\tau_\epsilon}\sum_{i=1}^{n}\sum_{j=1}^{J}\left\{Y_{ij} - m(X_{ij}) - (\overline{Y}_i - \overline{m}_i)\right\}^2 - \frac{J}{\tau_\alpha}\sum_{i=1}^{n}(\overline{Y}_i - \overline{m}_i)^2,$$

where $\overline{Y}_i = n^{-1}\sum_{j=1}^{J} Y_{ij}$ and $\overline{m}_i = n^{-1}\sum_{j=1}^{J} m(X_{ij})$. This "likelihood" is maximized at

$$\begin{aligned}
\widehat{\tau}_\alpha &= \frac{J}{n}\sum_{i=1}^{n}(\overline{Y}_i - \overline{m}_i)^2, \\
\widehat{\tau}_\epsilon &= \frac{1}{n(J-1)}\sum_{i=1}^{n}\sum_{j=1}^{J}\left\{Y_{ij} - m(X_{ij}) - (\overline{Y}_i - \overline{m}_i)\right\}^2,
\end{aligned}$$

when $\widehat{\tau}_\alpha > \widehat{\tau}_\epsilon$ and at

$$\widehat{\tau}_\alpha = \widehat{\tau}_\epsilon = \frac{1}{nJ}\sum_{i=1}^{n}\sum_{j=1}^{J}\{Y_{ij} - m(X_{ij})\}^2,$$

otherwise. Substituting a consistent estimator of $m(\cdot)$ yields consistent estimates of $(\sigma_\alpha^2, \sigma_\epsilon^2)$, and from results of Gutierrez & Carroll (1996) combined with (10), it can be shown that the resulting estimators have the same limit distribution as if $m(\cdot)$ actually were known.

However, as described below the covariance matrix of the residuals is *not* the same as if $m(\cdot)$ were known, and following the procedure used in many venues (e.g., Chambers & Hastie, 1992, pp. 368–369), we can adjust for the loss of degrees of freedom due to estimating $m(\cdot)$. In practice, we center the residuals at their mean, using $\underline{\mathbf{D}} = \underline{\mathbf{D}}_* - \underline{\mathbf{e}}_{nJ}^t\underline{\mathbf{D}}/(nJ)$, which has approximately mean zero and the covariance matrix $\underline{\mathbf{C}}(\sigma_\alpha^2, \sigma_\epsilon^2) = \sigma_\alpha^2\underline{\mathbf{C}}_1 + \sigma_\epsilon^2\underline{\mathbf{C}}_2$, where $\underline{\mathbf{C}}_1$ and $\underline{\mathbf{C}}_2$ are the known $(nJ)\times(nJ)$ matrices

$$\begin{aligned}
\underline{\mathbf{C}}_1 &= (\underline{\mathbf{I}} - \underline{\mathbf{S}})\mathrm{diag}(\underline{\mathbf{e}}_J\ \underline{\mathbf{e}}_J^t)(\underline{\mathbf{I}} - \underline{\mathbf{S}})^t; \\
\underline{\mathbf{C}}_2 &= (\underline{\mathbf{I}} - \underline{\mathbf{S}})(\underline{\mathbf{I}} - \underline{\mathbf{S}})^t.
\end{aligned}$$

In principle, we can still use normal-theory maximum likelihood (with covariance $\underline{\mathbf{C}}(\sigma_\alpha^2, \sigma_\epsilon^2) = \sigma_\alpha^2\underline{\mathbf{C}}_1$ $+\ \sigma_\epsilon^2\underline{\mathbf{C}}_2$) to estimate $(\sigma_\alpha^2, \sigma_\epsilon^2)$. However, the difficulty with maximum likelihood in this context is that the $(nJ)\times(nJ)$ matrix $\underline{\mathbf{C}}(\cdot)$ is impractical to invert. We consider two alternative methods of adjustment.

One approach is to make a restricted maximum likelihood (REML) style adjustment by substituting the estimate of $m(\cdot)$ into the estimating equations, taking their (approximate) expectations, subtracting these expectations from the original estimating equations and then solving the resulting (approximately) unbiased estimating equations. For the case that $\widehat{\tau}_\alpha > \widehat{\tau}_\epsilon$, after some considerable algebra, we obtain the approximately unbiased estimating equations

$$\begin{aligned} 0 &= \underline{\mathbf{D}}^t \underline{\mathbf{U}}\underline{\mathbf{U}}^t \underline{\mathbf{D}} - \tau_\epsilon w_1 - \frac{\tau_\alpha}{J} u_1, \\ 0 &= \underline{\mathbf{D}}^t \underline{\mathbf{D}} - \tau_\epsilon w_2 - \frac{\tau_\alpha}{J} u_2, \end{aligned}$$

where

$$\begin{aligned} w_r &= \text{trace}(\underline{\mathbf{W}}_r - \frac{1}{J}\underline{\mathbf{U}}\underline{\mathbf{U}}^t \underline{\mathbf{W}}_r), \\ u_r &= \text{trace}(\underline{\mathbf{U}}\underline{\mathbf{U}}^t \underline{\mathbf{W}}_r), \\ \underline{\mathbf{U}} &= \begin{pmatrix} \underline{\mathbf{e}}_J & \underline{\mathbf{0}} & \cdots & \underline{\mathbf{0}} \\ \underline{\mathbf{0}} & \underline{\mathbf{e}}_J & \cdots & \underline{\mathbf{0}} \\ . & \ldots & . & . \\ \underline{\mathbf{0}} & \ldots & \underline{\mathbf{0}} & \underline{\mathbf{e}}_J \end{pmatrix}, \\ \underline{\mathbf{W}}_1 &= (\underline{\mathbf{I}} - \underline{\mathbf{S}})^t \underline{\mathbf{U}}\underline{\mathbf{U}}^t (\underline{\mathbf{I}} - \underline{\mathbf{S}}), \\ \underline{\mathbf{W}}_2 &= (\underline{\mathbf{I}} - \underline{\mathbf{S}})^t (\underline{\mathbf{I}} - \underline{\mathbf{S}}). \end{aligned}$$

Solving these two equations, we obtain

$$\begin{aligned} \widehat{\tau}_\epsilon &= \frac{u_2 \underline{\mathbf{D}}^t \underline{\mathbf{U}}\underline{\mathbf{U}}^t \underline{\mathbf{D}} - u_1 \underline{\mathbf{D}}^t \underline{\mathbf{D}}}{w_1 u_2 - w_2 u_1}, \\ \widehat{\tau}_\alpha &= J \frac{w_1 \underline{\mathbf{D}}^t \underline{\mathbf{D}} - w_2 \underline{\mathbf{D}}^t \underline{\mathbf{U}}\underline{\mathbf{U}}^t \underline{\mathbf{D}}}{w_1 u_2 - w_2 u_1}. \end{aligned}$$

Alternatively, we can abandon the "likelihood" and employ a method of moments device. Let otrace$(\cdot)$ be the sum of the off-diagonal elements of a matrix. Then we can solve the two equations

$$\begin{aligned} \text{trace}(\underline{\mathbf{D}}\underline{\mathbf{D}}^t) &= \sigma_\alpha^2 \text{trace}(\underline{\mathbf{C}}_1) + \sigma_\epsilon^2 \text{trace}(\underline{\mathbf{C}}_2); \\ \text{otrace}(\underline{\mathbf{D}}\underline{\mathbf{D}}^t) &= \sigma_\alpha^2 \text{otrace}(\underline{\mathbf{C}}_1) + \sigma_\epsilon^2 \text{otrace}(\underline{\mathbf{C}}_2), \end{aligned}$$

so that

$$\begin{aligned} \widehat{\sigma}_\alpha^2 &= \frac{\text{otrace}(\underline{\mathbf{C}}_1)\text{trace}(\underline{\mathbf{D}}\ \underline{\mathbf{D}}^t) - \text{trace}(\underline{\mathbf{C}}_1)\text{otrace}(\underline{\mathbf{D}}\ \underline{\mathbf{D}}^t)}{\text{otrace}(\underline{\mathbf{C}}_1)\text{trace}(\underline{\mathbf{C}}_2) - \text{trace}(\underline{\mathbf{C}}_1)\text{otrace}(\underline{\mathbf{C}}_2)} \\ \widehat{\sigma}_\epsilon^2 &= \frac{\text{otrace}(\underline{\mathbf{C}}_2)\text{trace}(\underline{\mathbf{D}}\ \underline{\mathbf{D}}^t) - \text{trace}(\underline{\mathbf{C}}_2)\text{otrace}(\underline{\mathbf{D}}\ \underline{\mathbf{D}}^t)}{\text{otrace}(\underline{\mathbf{C}}_2)\text{trace}(\underline{\mathbf{C}}_1) - \text{trace}(\underline{\mathbf{C}}_2)\text{otrace}(\underline{\mathbf{C}}_1)}. \end{aligned}$$

These estimators can be shown to have the same limiting distribution as the method of moments estimators for known $m(\cdot)$, namely

$$\begin{aligned} \widehat{\sigma}_\alpha^2\{m(\cdot) \text{ known}\} &= \text{otrace}(\underline{\mathbf{E}}\ \underline{\mathbf{E}}^t)/\{nJ(nJ-1)\}; \\ \widehat{\sigma}_\epsilon^2\{m(\cdot) \text{ known}\} &= \{\text{trace}(\underline{\mathbf{E}}\ \underline{\mathbf{E}}^t)/(nJ)\} - \widehat{\sigma}_\alpha^2. \end{aligned}$$

6 DISCUSSION

{sec:discuss}

We have considered a number of different approaches (more than we have reported on here) to estimating the regression function when we have a simple dependence structure between observations. The simple pooled estimator which ignores the dependence structure performs very well asymptotically. Intuitively, this is because dependence is a global property of the error structure which (at least in the form we have examined) is not important to methods which act locally in the covariate space. Specifically, in the limit, local estimation methods are effectively dealing only with independent observations.

The performance of the pooled estimator raises the question of whether there is some method of local estimation which nonetheless exploits the dependence structure in such a way that it performs better than the pooled estimator. The quasi-likelihood estimator is very appealing for estimating the parametric component in a partially linear model and the general approach for estimating nonparametric components described by Carroll, Ruppert and Welsh (1996) suggests that the extension we have considered in this paper is well worth considering. We were surprised to find that quasi-likelihood estimation is asymptotically no better than pooled estimation. After trying a number of alternative approaches, we discovered that the two-step method has smaller asymptotic variance than the pooled estimator but does not necessarily have a lower asymptotic bias. The question of whether it is possible to construct an estimator with uniformly smaller asymptotic mean squared error than the pooled estimator remains open.

It is interesting to note that even if we were to assume a parametric form for the regression function, we would gain conflicting intuition into the problem of estimating the regression function in our problem. First notice that if we were to assume a constant regression function, then the maximum likelihood estimator (under Gaussianity) of the constant regression function is the sample mean which is, in this context, the pooled estimator. On the other hand, if we assume a linear regression function, the maximum likelihood estimator (under Gaussianity) of the linear regression function is the weighted least squares estimator which performs better than the least squares estimator which is, in this context, the pooled estimator. Thus the intuition we gain depends on which parametric model we consider.

REFERENCES

Carroll, R. J., Ruppert, D. & Welsh, A. (1996). Nonparametric estimation via local estimating equations, with applications to nutrition calibration. Preprint.

Carroll, R. J. & Wand, M. P. (1991). Semiparametric estimation in logistic measurement error models. *Journal of the Royal Statistical Society, Series B*, 53, 573–585.

Chambers, J. M. & Hastie, T. J. (1992). *Statistical Models in S*. Wadsworth & Brooks/Cole, Pacific Grove, CA.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87, 998-1004.

Gutierrez, R. G. & Carroll, R. J. (1996). Plug-in semiparametric estimating equations. *Biometrika*, to appear.

Ruppert, D., and Wand, M. P. (1994). Multivariate Locally Weighted Least Squares Regression, *Annals of Statistics*, 22, 1346–1370.

Severini, Th. A., and Staniswalis, J. G. (1994). Quasi-likelihood Estimation in Semiparametric Models, *Journal of the American Statistical Association*, 89, 501–511.

Appendix A PROOFS OF THEOREMS

{sec:app}

A.1 Proof of Theorem 1

{pr:pooled}

We first derive the results for the component estimator. From Fan (1992), Ruppert & Wand (1994) and Carroll, Ruppert & Welsh (1996), we have the results:

$$\begin{aligned}
&\mathrm{bias}\{\widehat{m}_{p,j}(x_0,h)\} \approx (1/2)h^2 m^{(2)}(x_0);\\
&\mathrm{var}\{\widehat{m}_{p,j}(x_0,h)\} \approx \gamma(\sigma_\alpha^2+\sigma_\epsilon^2)\left\{nhf_j(x_0)\right\}^{-1};\\
&\widehat{m}_{p,j}(x_0,h) - m(x_0,h) - (h^2/2)m^{(2)}(x_0) \approx \{nf_j(x_0)\}^{-1}\sum_{i=1}^{n}\{Y_{ij}-m(X_{ij})\}K_h(X_{ij}-x_0). \quad (10)
\end{aligned}$$

{eq:qc2}

The last step is implicit in the first two papers and explicit in the third. It is easily seen from (10) that for $j \neq k$, $\mathrm{cov}\{\widehat{m}_{p,j}(x_0,h), \widehat{m}_{p,k}(x_0,h)\} = O(n^{-1})$, and hence for asymptotics arguments, the component estimators $\widehat{m}_{p,j}(x_0,h)$ are independent.

It thus follows that

$$\begin{aligned}
\mathrm{bias}\{\widehat{m}_W(x_0,\underline{\mathbf{h}},\underline{\mathbf{c}})\} &\approx (1/2)m^{(2)}(x_0)\sum_{j=1}^{J} c_j h_j^2 = \sum_{j=1}^{J} c_j b_j(x_0,h_j);\\
\mathrm{var}\{\widehat{m}_W(x_0,\underline{\mathbf{h}},\underline{\mathbf{c}})\} &\approx \gamma(\sigma_\alpha^2+\sigma_\epsilon^2)n^{-1}\sum_{j=1}^{J} c_j^2\left\{h_j f_j(x_0)\right\}^{-1} = \sum_{j=1}^{J} c_j^2 v_j(x_0,h_j).
\end{aligned}$$

The individual component bias functions are $b_j(x_0,h_j)$ and the individual component variance functions are $v_j(x_0,h_j)$. The problem becomes to minimize (in $\underline{\mathbf{h}}$ and $\underline{\mathbf{c}}$) the function

$$\mathrm{mse}_W(x_0,\underline{\mathbf{h}},\underline{\mathbf{c}}) \approx \left\{\sum_{j=1}^{J} c_j b_j(x_0,h_j)\right\}^2 + \sum_{j=1}^{J} c_j^2 v_j(x_0,h_j)$$

subject to $\underline{\mathbf{e}}_J^t \underline{\mathbf{c}} = 1$. As shown in Appendix A.2, the minimization problem is solved by choosing a common bandwidth

$$h_{\text{opt}} = \left(\frac{\gamma(\sigma_\alpha^2 + \sigma_\epsilon^2)}{\{m^{(2)}(x_0)\}^2 n \sum_{j=1}^J f_j(x_0)} \right)^{1/5}$$

and weights

$$c_{\text{opt},j} = f_j(x_0) \left\{ \sum_{k=1}^J f_k(x_0) \right\}^{-1} \quad \text{for } j = 1, \ldots, J.$$

The asymptotic mean squared error at this optimal solution is

$$\text{mse}_{\text{opt},W} = 5/4 \left\{ m^{(2)}(x_0) \right\}^{2/5} \left\{ \frac{\gamma(\sigma_\alpha^2 + \sigma_\epsilon^2)}{n \sum_{j=1}^J f_j(x_0)} \right\}^{4/5}$$

which is (6).

We now turn to the pooled estimator. We proceed more generally than for local linear regression, obtaining results for local polynomial regression of order p with p odd. Let $\beta_j = h^j m^{(j)}(x_0)/j!$, and define $\underline{\mathcal{B}} = (\beta_0, ..., \beta_p)^t$. Let $G_p(v) = (1, v, v^2, ..., v^p)^t$. Then,

$$\widehat{\underline{\mathcal{B}}} - \underline{\mathcal{B}} \quad = \quad A^{-1}(n, J, h, x_0) B(n, J, h, x_0),$$

where

$$\begin{aligned} A(n, J, h, x_0) \quad &= \quad (nJ)^{-1} \sum_{i=1}^n \sum_{j=1}^J K_h(X_{ij} - x_0) G_p\{(X_{ij} - x_0)/h\} G_p^t\{(X_{ij} - x_0)/h\}; \\ B(n, J, h, x_0) \quad &= \quad (nJ)^{-1} \sum_{i=1}^n \sum_{j=1}^J K_h(X_{ij} - x_0) G_p\{(X_{ij} - x_0)/h\} \left\{ Y_{ij} - \sum_{k=1}^p m^{(k)}(x_0)(X_{ij} - x_0)^k/k! \right\}. \end{aligned}$$

Let $\mu(\ell) = \int z^\ell K(z) dz$ and $D_p(\mu)$ be the $(p+1) \times (p+1)$ matrix with (j,k)th element $\mu(j+k-2)$. Let $\gamma(\ell) = \int z^\ell K^2(z) dz$ so that $\gamma = \gamma(0)$, and let $D_p(\gamma)$ be the $(p+1) \times (p+1)$ matrix with (j,k)th element $\gamma(j+k-2)$. Direct calculations (keeping in mind that p is odd) show that

$$\begin{aligned} A(n, J, h, x_0) \quad &= \quad \{1 + o_p(1)\} s(x_0) D_p(\mu); \\ \text{cov}\{B(n, J, h, x_0)\} \quad &\approx \quad (nhJ)^{-1} s(x_0)(\sigma_\alpha^2 + \sigma_\epsilon^2) D_p(\gamma); \\ E\{B(n, J, h, x_0)\} \quad &\approx \quad \{s(x_0) m^{(p+1)}(x_0) h^{p+1}/(p+1)!\} \{\mu(p+1), ..., \mu(2p+1)\}^t. \end{aligned}$$

Thus, for p odd we have

$$\begin{aligned} \text{bias}\{\widehat{m}_{p,pool}(x_0, h)\} \quad &\approx \quad \left\{ m^{(p+1)}(x_0) h^{p+1}/(p+1)! \right\} (1, 0, ..., 0) D_p^{-1}(\mu) \{\mu(p+1), ..., \mu(2p+1)\}^t; \\ \text{var}\{\widehat{m}_{p,pool}(x_0, h)\} \quad &\approx \quad \{nhJs(x_0)\} (\sigma_\alpha^2 + \sigma_\epsilon^2)(1, 0, ..., 0) D_p(\gamma) D_p^{-1}(\mu) D_p^{-1}(\mu)(1, 0, ..., 0)^t. \end{aligned}$$

In the special case $p = 1$, these results reduce to (3)–(4), thus completing the proof.

A.2 Optimal Bandwidths and Weights for the Component Estimator

{pr:optbw}

To determine the optimal bandwidths and weights for the component estimator we have to minimize (in $\underline{\mathbf{h}}$, $\underline{\mathbf{c}}$, and λ)

$$\left\{\sum_{j=1}^{J} c_j b_j(x_0, h_j)\right\}^2 + \sum_{j=1}^{J} c_j^2 v_j(x_0, h_j) + \lambda \underline{\mathbf{e}}_J^t \underline{\mathbf{c}}\,,$$

where λ is the Lagrange multiplier. To simplify the notation, define the matrix $\underline{\mathbf{H}} = \mathrm{diag}(\underline{\mathbf{h}})$ and the vector $\underline{\mathbf{f}}(x_0) = \{f_1(x_0), \ldots, f_J(x_0)\}^t$. Taking partial derivatives, the optimal parameters satisfy the implicit equations

$$\{m^{(2)}(x_0)\}^2 \underline{\mathbf{c}}^t \underline{\mathbf{H}}^2 \underline{\mathbf{e}}_J\ \underline{\mathbf{H}}^3 \underline{\mathbf{f}}(x_0) - \gamma(\sigma_\alpha^2 + \sigma_\epsilon^2) n^{-1} \underline{\mathbf{c}} = 0 \quad (11) \quad \{eq:pd1\}$$

$$2^{-1}\{m^{(2)}(x_0)\}^2 \underline{\mathbf{c}}^t \underline{\mathbf{H}}^2 \underline{\mathbf{e}}_J\ \underline{\mathbf{H}}^3 \underline{\mathbf{f}}(x_0) + 2\gamma(\sigma_\alpha^2 + \sigma_\epsilon^2) n^{-1} \underline{\mathbf{c}} + \lambda \underline{\mathbf{H}}\underline{\mathbf{f}}(x_0) = 0 \quad (12) \quad \{eq:pd2\}$$

$$\underline{\mathbf{c}}^t \underline{\mathbf{e}}_J - 1 = 0. \quad (13) \quad \{eq:pd3\}$$

The difference between (11) and two times (12) results in

$$5\gamma(\sigma_\alpha^2 + \sigma_\epsilon^2) n^{-1} \underline{\mathbf{c}} + 2\lambda \underline{\mathbf{H}}\underline{\mathbf{f}}(x_0) = 0. \quad (14) \quad \{eq:pdd12\}$$

Multiplying this equation by $\underline{\mathbf{e}}_J^t$ and using (13), we obtain

$$\lambda = -5\gamma(\sigma_\alpha^2 + \sigma_\epsilon^2) \left\{2n \underline{\mathbf{e}}_J^t \underline{\mathbf{H}}\underline{\mathbf{f}}(x_0)\right\}^{-1}, \quad (15) \quad \{eq:lambda\}$$

and thus by substituting λ in (14) by the previous expression we obtain

$$\underline{\mathbf{c}} = \underline{\mathbf{H}}\underline{\mathbf{f}}(x_0) \left\{\underline{\mathbf{e}}_J^t \underline{\mathbf{H}}\underline{\mathbf{f}}(x_0)\right\}^{-1}. \quad (16) \quad \{eq:solc\}$$

We now turn to determining the optimal bandwidths $\underline{\mathbf{h}}$. Substitute λ and $\underline{\mathbf{c}}$ in (12) by (15) and (16). This results in

$$\left[\{m^{(2)}(x_0)\}^2 \underline{\mathbf{f}}(x_0)^t \underline{\mathbf{H}}^3 \underline{\mathbf{e}}_J \underline{\mathbf{H}}^2 - \gamma(\sigma_\alpha^2 + \sigma_\epsilon^2) n^{-1} \underline{\mathbf{I}}\right] \underline{\mathbf{H}}\underline{\mathbf{f}}(x_0) = 0,$$

which is satisfied if

$$\{m^{(2)}(x_0)\}^2 \underline{\mathbf{f}}(x_0)^t \underline{\mathbf{H}}^3 \underline{\mathbf{e}}_J\ \underline{\mathbf{H}}^2 - \gamma(\sigma_\alpha^2 + \sigma_\epsilon^2) n^{-1} \underline{\mathbf{I}} = 0.$$

We can conclude from this equation that h_j is constant for all $j = 1, \ldots, J$, and thus

$$h_{\mathrm{opt}}^5 = \frac{\gamma(\sigma_\alpha^2 + \sigma_\epsilon^2)}{\{m^{(2)}(x_0)\}^2\, n \sum_{j=1}^{J} f_j(x_0)}.$$

Consequently we obtain an optimal c_j of

$$c_{\mathrm{opt},j} = f_j(x_0) \left\{\sum_{k=1}^{J} f_k(x_0)\right\}^{-1},$$

as claimed.

A.3 Proof of Theorem 2

{pr:qle}

Here we give a brief derivation of the asymptotic bias and variance formulae for the quasi-likelihood estimator. We start by obtaining some general results for the local polynomial estimator. Then we obtain results for the kernel (local average), local linear, and local quadratic estimator.

A useful simplification is to let the unknown parameters be $\beta_q = h^q m^{(q)}(x_0)/q!$. Thus, $\underline{\mathbf{m}}(\underline{\mathbf{X}})$ is approximated by $\{\underline{\mathbf{G}}_{p,h}(\underline{\mathbf{X}} - x_0\underline{\mathbf{e}}_J)\}^t \underline{\beta}$, where $\underline{\mathbf{G}}_{p,h}(\underline{\mathbf{X}} - x_0\underline{\mathbf{e}}_J)$ is the $(p+1)\times J$ matrix with $(k+1)$th row $[\{(X_1 - x_0)/h\}^k, \{(X_2 - x_0)/h\}^k, \ldots, \{(X_J - x_0)/h\}^k]$. Define

$$\underline{\mathbf{L}}_n(\underline{\alpha}) = n^{-1}\sum_{i=1}^{n} \underline{\mathbf{G}}_{p,h}(\underline{\mathbf{X}}_i - x_0\underline{\mathbf{e}}_J)\,\underline{\mathbf{V}}\underline{\mathbf{K}}_h(\underline{\mathbf{X}}_i - x_0\underline{\mathbf{e}}_J)\,\left[\underline{\mathbf{Y}}_i - \left\{\underline{\mathbf{G}}_{p,h}(\underline{\mathbf{X}}_i - x_0\underline{\mathbf{e}}_J)\right\}^t \underline{\alpha}\right],$$

where $\underline{\mathbf{K}}_h(\underline{\mathbf{X}}_i - x_0\underline{\mathbf{e}}_J) = \mathrm{diag}\{K_h(X_{i1}-x_0), \ldots, K_h(X_{iJ}-x_0)\}$. Then the quasi-likelihood estimator (at x_0) solves $\underline{\mathbf{0}} = \underline{\mathbf{L}}_n(\underline{\widehat{\beta}})$ and hence

$$\underline{\widehat{\beta}} - \underline{\beta} = (\underline{\mathbf{B}}_n^*)^{-1}\,\underline{\mathbf{L}}_n(\underline{\beta}),$$

where $\underline{\mathbf{B}}_n^*$ is the $(p+1)\times(p+1)$ matrix

$$\underline{\mathbf{B}}_n^* = n^{-1}\sum_{i=1}^{n} \underline{\mathbf{G}}_{p,h}(\underline{\mathbf{X}}_i - x_0\underline{\mathbf{e}}_J)\,\underline{\mathbf{V}}\underline{\mathbf{K}}_h(\underline{\mathbf{X}}_i - x_0\underline{\mathbf{e}}_J)\,\{\underline{\mathbf{G}}_{p,h}(\underline{\mathbf{X}}_i - x_0\underline{\mathbf{e}}_J)\}^t.$$

Let $v_{j,k}$ be the elements of $\underline{\mathbf{V}}$. Then the elements of the matrix $\underline{\mathbf{B}}_n^*$ are

$$(\underline{\mathbf{B}}_n^*)_{r,s} = \sum_{k=1}^{J}\sum_{j=1}^{J} v_{j,k} A_{n,j,k}(r-1, s-1),$$

where

$$A_{n,j,k}(r,s) = n^{-1}\sum_{i=1}^{n}\left(\frac{X_{ij}-x_0}{h}\right)^r K_h(X_{ik}-x_0)\left(\frac{X_{ik}-x_0}{h}\right)^s.$$

It is easily seen that

$$A_{n,k,k}(r,s) \sim \int z^{r+s} K(z) f_k(x_0 + z\,h)\,dz$$

and

$$A_{n,j,k}(r,s) \sim h^{-r}\int\int (x_j - x_0)^r z^s K(z) f_{j,k}(x_j,\, x_0 + zh)\,dx_j\,dz$$

for $j \neq k$. Thus, the elements of $\underline{\mathbf{B}}^*$, the limit of $\underline{\mathbf{B}}_n^*$ as $n \to \infty$, are

$$(\underline{\mathbf{B}}^*)_{r,s} \;=\; \sum_{\ell=0}^{m}\left\{h^{\ell+1-r}\mu(\ell+s-1)B(r-1,\ell)\right\} + \mu(m+s)\mathcal{O}_P(h^{m+2-r}), \qquad (17)$$ {eq:asB}

where

$$
\begin{aligned}
B(0,\ell) &= \sum_{k=1}^{J}\left\{\frac{f_k^{(\ell)}(x_0)}{\ell!}\sum_{j=1}^{J} v_{j,k}\right\} \\
B(r,\ell) &= \sum_{k=1}^{J}\sum_{j\neq k}^{J}\frac{v_{j,k}}{\ell!}\int (x_j-x_0)^r \frac{\partial^\ell}{(\partial x_0)^\ell} f_{j,k}(x_j,x_0)dx_j + 1_{[\ell\geq r]}\sum_{k=1}^{J}\frac{v_{k,k}f_k^{(\ell-r)}(x_0)}{(\ell-r)!}
\end{aligned}
$$

for $r>0$ and for $\ell\geq 0$.

Furthermore, note that since $E\{\underline{\mathbf{Y}}-\underline{\mathbf{m}}(\underline{\mathbf{X}})|\underline{\mathbf{X}}\}=0$,

$$
E\{\underline{\mathbf{L}}_n(\underline{\beta})\} = E_{\underline{\mathbf{X}}}\left(\underline{\mathbf{G}}_{p,h}(\underline{\mathbf{X}}-x_0\underline{\mathbf{e}}_J)\,\underline{\mathbf{V}}\underline{\mathbf{K}}_h(\underline{\mathbf{X}}-x_0\underline{\mathbf{e}}_J)\,\left[\underline{\mathbf{m}}(\underline{\mathbf{X}})-\left\{\underline{\mathbf{G}}_{p,h}(\underline{\mathbf{X}}-x_0\underline{\mathbf{e}}_J)\right\}^t\underline{\beta}\right]\right).
$$

But Taylor's theorem implies

$$
\begin{aligned}
\underline{\mathbf{m}}(\underline{\mathbf{X}})-\left\{\underline{\mathbf{G}}_{p,h}(\underline{\mathbf{X}}-x_0\underline{\mathbf{e}}_J)\right\}^t\underline{\beta} &= \beta_{p+1}\{(\underline{\mathbf{X}}-x_0\underline{\mathbf{e}}_J)/h\}^{p+1}+\beta_{p+2}\{(\underline{\mathbf{X}}-x_0\underline{\mathbf{e}}_J)/h\}^{p+2}+ \\
&\quad \left[\mathcal{O}(h^{p+3}\{(\mathbf{X}_1-x_0\underline{\mathbf{e}}_J)/h\}^{p+3}),\ldots,\mathcal{O}(h^{p+3}\{(\mathbf{X}_1-x_0\underline{\mathbf{e}}_J)/h\}^{p+3})\right]^t
\end{aligned}
$$

and thus

$$
\begin{aligned}
\left[E\{\underline{\mathbf{L}}_n(\underline{\beta})\}\right]_r &\approx \frac{h^{p+1}m^{(p+1)}(x_0)}{(p+1)!}(\underline{\mathbf{B}}^*)_{r,p+2}+\frac{h^{p+2}m^{(p+2)}(x_0)}{(p+2)!}(\underline{\mathbf{B}}^*)_{r,p+3} \\
&\approx \frac{h^{p+1}m^{(p+1)}(x_0)}{(p+1)!}\left\{h^{1-r}\mu(p+1)B(r-1,0)+h^{2-r}\mu(p+2)B(r-1,1)\right\} \\
&\quad + \frac{h^{p+2}m^{(p+2)}(x_0)}{(p+2)!}h^{1-r}\mu(p+2)B(r-1,0). \qquad (18)
\end{aligned}
$$

{eq:gEL}

for $r=1,2,\ldots,p+1$. The asymptotic bias is then

$$
\text{bias}\left(\widehat{m}_{1,qle}(x_0,h)\right)\approx(1,0,0,\ldots)\left(\underline{\mathbf{B}}^*\right)^{-1}E\{\underline{\mathbf{L}}_n(\underline{\beta})\} \qquad (19)
$$

{eq:gbias}

The covariance matrix of $\underline{\mathbf{Y}}$ given $\underline{\mathbf{X}}$ is $\underline{\Sigma}=(\sigma_{j,k})$. Thus the covariance of $\underline{\mathbf{L}}_n(\underline{\beta})$ is $\underline{\mathbf{C}}_{\underline{\mathbf{L}}_n}$, with elements given by

$$
\begin{aligned}
(\underline{\mathbf{C}}_{\underline{\mathbf{L}}_n})_{r,s} &= \frac{1}{n}E_{\underline{\mathbf{X}}}\left\{\sum_{j=1}^{J}\sum_{k=1}^{J}\sum_{\ell=1}^{J}\sum_{m=1}^{J} v_{\ell,j}\sigma_{j,k}v_{k,m}\right. \\
&\quad \left.\times\; K_h(X_{1j}-x_0)K_h(X_{1k}-x_0)\left(\frac{X_{1\ell}-x_0}{h}\right)^{r-1}\left(\frac{X_{1m}-x_0}{h}\right)^{s-1}\right\}.
\end{aligned}
$$

Let $v_{k\cdot}=\sum_{j=1}^{J}v_{k,j}$. Then direct, but lengthy, calculations show that

$$
\begin{aligned}
(\underline{\mathbf{C}}_{\underline{\mathbf{L}}_n})_{1,1} &\approx \gamma(0)n^{-1}h^{-1}\sum_{k=1}^{J}(v_{k\cdot})^2\sigma_{k,k}f_k(x_0) \qquad (20) \\
(\underline{\mathbf{C}}_{\underline{\mathbf{L}}_n})_{1,s} &\approx \gamma(0)n^{-1}h^{-s}\sum_{k=1}^{J}\left\{v_{k\cdot}\sigma_{k,k}\sum_{m\neq k}^{J}v_{k,m}\int(x_m-x_0)^{s-1}f_{k,m}(x_0,x_m)dx_m\right\} \qquad (21)
\end{aligned}
$$

{eq:cov11} {eq:cov1.}

$$(\underline{\mathbf{C}}_{\underline{\mathbf{L}}_n})_{r,s} \approx \gamma(0) n^{-1} h^{-r-s+1} \sum_{k=1}^{J} \Big\{ \sigma_{k,k} \sum_{\ell \neq k}^{J} \sum_{m \neq k}^{J} v_{\ell,k} v_{k,m} \times \int (x_\ell - x_0)^{r-1} (x_m - x_0)^{s-1} f_{k,\ell,m}(x_0, x_\ell, x_m) dx_\ell dx_m \Big\} \quad (22) \quad \{eq{:}covrs\}$$

for $r \geq s$ and for $s > 1$ and thus, because the covariance matrix is symmetric, we have a first order approximation of the asymptotic covariance matrix. The variance of the quasi-likelihood estimator is then

$$\mathrm{var}\{\widehat{m}_{2,qle}(x_0,h)\} \approx (1,0,0,\ldots)(\underline{\mathbf{B}}^*)^{-1} \underline{\mathbf{C}}_{\underline{\mathbf{L}}_n} \left\{ (\underline{\mathbf{B}}^*)^{-1} \right\}^t (1,0,0,\ldots)^t.$$

A.3.1 Kernel (Local Average) Estimation

¿From the previous calculation it is easy to determine the bias and variance of the kernel estimator ($p = 0$). Using (17), (18), (19 and (20), respectively, the bias is

$$\begin{aligned} \{(\underline{\mathbf{B}}^*)_{1,1}\}^{-1} E\{\underline{\mathbf{L}}_n(\underline{\beta})\} &\approx B(0,0)^{-1} \left\{ h m^{(1)}(x_0) h B(0,1) + h^2 m^{(2)}(x_0)/2\, B(0,0) \right\} \\ &= h^2 \left\{ m^{(1)}(x_0) \frac{\sum_{k=1}^{J} v_{k\cdot} f_k^{(1)}(x_0)}{\sum_{k=1}^{J} v_{k\cdot} f_k(x_0)} + m^{(2)}(x_0)/2 \right\}. \end{aligned}$$

and the asymptotic variance is

$$\mathrm{var}\{\widehat{m}_{1,qle}\} = \frac{(\underline{\mathbf{C}}_{\underline{\mathbf{L}}_n})_{1,1}}{(\underline{\mathbf{B}}^*)_{1,1}^2} \approx \frac{\gamma(0) \left\{ \sum_{k=1}^{J} (v_{k\cdot})^2 \sigma_{k,k} f_k(x_0) \right\}}{nh \left(\sum_{k=1}^{J} v_{k\cdot} f_k(x_0) \right)^2}$$

as claimed. The bias and variance for the variance component model are easily determined from the above results since $v_{k\cdot}$ and $\sigma_{k,k}$ are constant (in k).

A.3.2 Local Linear Estimation

For the local linear estimator ($p = 1$), the calculation of $\underline{\mathbf{B}}^*$ results in

$$\underline{\mathbf{B}}^* \approx \begin{bmatrix} B(0,0) & h\, B(0,1) \\ h^{-1} B(1,0) & B(1,1) \end{bmatrix}$$

since $K(\cdot)$ is symmetric ($\mu(r) = 0$ if r is odd) and $\mu(r) = 1$ for $r = 0, 2$. Note that $(\underline{\mathbf{B}}^*)_{2,1}$ tends to infinity as h goes to zero. Therefore the limit, $h \to 0$, is taken after $(\underline{\mathbf{B}}^*)^{-1} \underline{\mathbf{Y}}^*$ has been calculated. The determinant of $\underline{\mathbf{B}}^*$ is

$$\det(\underline{\mathbf{B}}^*) \approx B(0,0)B(1,1) - B(0,1)B(1,0).$$

A direct calculation of $E\{\underline{\mathbf{L}}_n(\underline{\beta})\}$ for $p = 1$ yields

$$E\{\underline{\mathbf{L}}_n(\underline{\beta})\} \approx h^2\, m^{(2)}(x_0)/2 \begin{bmatrix} B(0,0) \\ h^{-1} B(1,0) \end{bmatrix}$$

using (18). Thus, the bias (19) is

$$\text{bias}\{\widehat{m}_{1,qle}(x_0,h)\} \approx (1/2)h^2 m^{(2)}(x_0)$$

as claimed.

The variance is

$$\begin{aligned}
\text{var}\{\widehat{m}_{1,qle}(x_0,h)\} &= (1,0)\,(\underline{\mathbf{B}}^*)^{-1}\mathbf{C}_{\underline{\mathbf{L}}_n}\left\{(\underline{\mathbf{B}}^*)^{-1}\right\}^t\,(1,0)^t \\
&\approx \frac{B(1,1)^2(\mathbf{C}_{\underline{\mathbf{L}}_n})_{1,1} - 2hB(0,1)B(1,1)(\mathbf{C}_{\underline{\mathbf{L}}_n})_{1,2} + h^2B(0,1)^2(\mathbf{C}_{\underline{\mathbf{L}}_n})_{2,2}}{\{B(0,0)B(1,1)-B(0,1)B(1,0)\}^2}
\end{aligned}$$

This expression is of order $\gamma(0)/(nh)$ times a quantity which is nonzero in general.

For the variance component model the off-diagonal elements of $V=\Sigma^{-1}$, the diagonal elements of V and of Σ, and $v_{k\cdot}$ are constant called v_o, v_d, σ_d^2, and $v.$, respectively. To simplify the calculation further, we assume the X_j's are independent with a common marginal density function f. Then the key quantities reduce to

$$\begin{aligned}
B(0,\ell) &= v.J\,f^{(\ell)}(x_0)/(\ell!) \\
B(r,\ell) &= v_oJ(J-1)f^{(\ell)}(x_0)/(\ell!)\,E\left\{(X-x_0)^r\right\} + 1_{[\ell\geq r]}v_dJ\,f^{(\ell-r)}(x_0)/\{(\ell-r)!\} \\
(\mathbf{C}_{\underline{\mathbf{L}}_n})_{1,1} &\approx \gamma(0)n^{-1}h^{-1}v_\cdot^2\sigma_d^2J\,f(x_0) \\
(\mathbf{C}_{\underline{\mathbf{L}}_n})_{1,s} &\approx \gamma(0)n^{-1}h^{-s}v.v_o\sigma_d^2J(J-1)\,f(x_0)E\left\{(X-x_0)^s\right\} \\
(\mathbf{C}_{\underline{\mathbf{L}}_n})_{r,s} &\approx \gamma(0)n^{-1}h^{-r-s+1}v_o^2\sigma_d^2J(J-1)\,f(x_0)\left[E\left\{(X-x_0)^{r+s-2}\right\}\right. \\
&\qquad \left. +\ (J-2)E\left\{(X-x_0)^{r-1}\right\}E\left\{(X-x_0)^{s-1}\right\}\right].
\end{aligned}$$

Thus the determinant of $\underline{\mathbf{B}}^*$ is $\det(\underline{\mathbf{B}}^*)\approx v.v_d\{J\,f(x_0)\}^2$ and the above expression of the variance results in

$$\text{var}\{\widehat{m}_{1,qle}(x_0,h)\} \approx \frac{\gamma}{nh}\frac{\sigma_d^2}{Jf(x_0)}\left\{1+\left(\frac{v_o}{v_d}\frac{f^{(1)}(x_0)}{f(x_0)}\right)^2(J-1)\sigma_X^2\right\},$$

where σ_X^2 is the variance of X. Remembering that $\sigma_d^2=\sigma_\alpha^2+\sigma_\epsilon^2$, $v_d=(1-\eta/J)/\sigma_\epsilon^2$ and $v_d = -\eta/(J\sigma_\epsilon^2)$ yields $v_o/v_d=-\eta/(J-\eta)$. This completes the proof.

A.3.3 Local Quadratic Estimation

For the local quadratic estimation ($p=2$), the first two nonzero terms of the approximation of $\underline{\mathbf{B}}^*$ are

$$\underline{\mathbf{B}}^* \approx \left\{\begin{array}{ccc} B(0,0) & h\,B(0,1) & B(0,0) \\ h^{-1}B(1,0) & B(1,1) & h^{-1}B(1,0) \\ h^{-2}B(2,0) & h^{-1}B(2,1) & h^{-2}B(2,0)\end{array}\right\} + \left\{\begin{array}{ccc} h^2B(0,2) & h^3\mu(4)B(0,3) & h^2\mu(4)B(0,2) \\ h\,B(1,2) & h^2\mu(4)B(1,3) & h\mu(4)B(1,2) \\ B(2,2) & h^1\mu(4)B(2,3) & \mu(4)B(2,2)\end{array}\right\}.$$

Matrix algebra shows that the determinant of this sum of singular matrices is

$$
\begin{aligned}
&= h^2\{\mu(4)-1\}\ \det\left\{\begin{array}{ccc} B(0,0) & h\,B(0,1)+h^3\mu(4)B(0,3) & \mu(4)B(0,2) \\ h^{-1}B(1,0) & B(1,1)+h^2\mu(4)B(1,3) & h^{-1}\mu(4)B(1,2) \\ h^{-2}B(2,0) & h^{-1}B(2,1)+h^1\mu(4)B(2,3) & h^{-2}\mu(4)B(2,2) \end{array}\right\} \\
&\overset{h\to 0}{\longrightarrow} \{\mu(4)-1\}\,S_D,
\end{aligned}
$$

where

$$
\begin{aligned}
S_D &= B(0,0)B(1,1)B(2,2) - B(0,0)B(2,1)B(1,2) - B(1,0)B(0,1)B(2,2) \\
&\quad + B(1,0)B(2,1)B(0,2) + B(2,0)B(0,1)B(1,2) - B(2,0)B(0,2)B(1,1), \qquad (23)
\end{aligned}
$$

{eq:SD}

assuming $\mu(4) \neq 1$ and $S_D \neq 0$. (Otherwise the determinant is of order h^2.)

To calculate (19) we again need the first two terms of the approximation of $\underline{\mathbf{B}}^*$ since using only the first terms in the multiplication results in all terms cancelling out. However, even this is not enough: we also need terms of order h^{p+4} in the approximation of $\underline{\mathbf{L}}_n(\underline{\beta})$ in (18). Note that because the first terms in the approximation of $(\underline{\mathbf{B}}^*)_{r,6}$ (for $m^{(5)}$) and $(\underline{\mathbf{B}}^*)_{r,7}$ (for $m^{(7)}$) are proportional to the first terms in $(\underline{\mathbf{B}}^*)_{r,4}$ (for $m^{(3)}$) and $(\underline{\mathbf{B}}^*)_{r,5}$ (for $m^{(4)}$), respectively, these terms cancel out as the latter have done. Then direct calculations show that the asymptotic bias is

$$
\text{bias}\{\widehat{m}_{2,qle}(x_0,h)\} \approx h^4 \frac{\mu(6)-\mu^2(4)}{\mu(4)-1}\left\{\frac{m^{(3)}(x_0)}{3!}\frac{S_N}{S_D} - \frac{m^{(4)}(x_0)}{4!}\right\},
$$

where

$$
\begin{aligned}
S_N &= B(0,3)B(1,1)B(2,0) - B(0,3)B(2,1)B(1,0) - B(1,3)B(0,1)B(2,0) \\
&\quad + B(1,3)B(2,1)B(0,0) + B(2,3)B(0,1)B(1,0) - B(2,3)B(1,1)B(0,0). \qquad (24)
\end{aligned}
$$

{eq:SN}

The first nonzero orders of the elements of $\underline{\mathbf{C}}_{\underline{\mathbf{L}}_n}$ are

$$
\frac{1}{n}\begin{bmatrix} \mathcal{O}(h^{-1}) & \mathcal{O}(h^{-2}) & \mathcal{O}(h^{-3}) \\ \mathcal{O}(h^{-2}) & \mathcal{O}(h^{-3}) & \mathcal{O}(h^{-4}) \\ \mathcal{O}(h^{-3}) & \mathcal{O}(h^{-4}) & \mathcal{O}(h^{-5}) \end{bmatrix}.
$$

The order of the terms in the first row of $(\underline{\mathbf{B}}^*)^{-1}$ is $\{\mathcal{O}(h^{-2}), \mathcal{O}(h^{-1}), \mathcal{O}(1)\}$. Thus the leading order of the variance is

$$
\text{var}\{\widehat{m}_{2,qle}(x_0,h)\} \;=\; (1,0,0)\,\text{cov}\{(\underline{\mathbf{B}}^*)^{-1}\underline{\mathbf{Y}}^*\}\,(1,0,0)^t = \mathcal{O}_P(n^{-1}h^{-5})
$$

For the variance component model with common marginal distribution, a direct calculation results in

$$
\begin{aligned}
\det(\underline{\mathbf{B}}^*) &\overset{h\to 0}{\longrightarrow} v.v_d^2\{J\,f(x_0)\}^3\{\mu(4)-1\} \\
\text{var}\{\widehat{m}_{2,qle}(x_0,h)\} &\approx \frac{\gamma}{nh^5}\frac{\sigma_d^2 v_o^2}{J\,f(x_0)\,v_d^2}\frac{(J-1)}{\{\mu(4)-1\}^2}\left\{\int (x-x_0)^4 f(x)\,dx - \left(\sigma_{\underline{\mathbf{X}}}^2\right)^2\right\}.
\end{aligned}
$$

Replacing σ_d^2, v_d, and v_o by their actual values yields the result in the theorem.

A.4 Proof of Theorem 3

{pr:mqle}

If we set the off-diagonal elements of $\underline{\mathbf{V}}$ to zero ($v_{j,k} = 0$ for $j \neq k$), then the key quantities reduce to

$$\begin{aligned}
(\underline{\mathbf{B}}^*)_{r,s} &= \sum_{\ell=0} \left\{ \mu(r+s+\ell-2)h^\ell \sum_{j=1}^J v_{j,j} f_j^{(\ell)}(x_0) \right\} \approx \mu(r+s-2) \sum_{j=1}^J v_{j,j} f_j(x_0) \\
\left[E\{\underline{\mathbf{L}}_n(\underline{\beta})\} \right]_r &\approx \mu(r+p)h^{p+1} \frac{m^{(p+1)}(x_0)}{(p+1)!} \sum_{j=1}^J v_{j,j} f_j(x_0) \\
&\quad + \mu(r+p+1)h^{p+2} \sum_{j=1}^J v_{j,j} \left\{ \frac{m^{(p+1)}(x_0)}{(p+1)!} f_j^{(1)}(x_0) + \frac{m^{(p+2)}(x_0)}{(p+2)!} f_j(x_0) \right\} \\
(\underline{\mathbf{C}}_{\underline{\mathbf{L}}_n})_{r,s} &\approx \gamma(r+s-2) n^{-1} h^{-1} \sum_{j=1}^J v_{j,j}^2 \sigma_{j,j} f_j(x_0)
\end{aligned}$$

Direct calculations then yield the results given in the theorem 2.

A.5 Proof of Theorem 4

{pr:comb}

The two-step estimator is just

$$\begin{aligned}
\widehat{m}_C(x_0) &= (1,0) \left\{ n^{-1} \sum_{i=1}^n \sum_{j=1}^J \begin{pmatrix} 1 \\ (X_{ij}-x_0)/h \end{pmatrix} \begin{pmatrix} 1 \\ (X_{ij}-x_0)/h \end{pmatrix}^t K_h(X_{ij}-x_0) \right\}^{-1} \\
&\quad \times \left\{ n^{-1} \sum_{i=1}^n \sum_{j=1}^J \begin{pmatrix} 1 \\ (X_{ij}-x_0)/h \end{pmatrix} Z_{ij} K_h(X_{ij}-x_0) \right\}. \qquad (25)
\end{aligned}$$

{eq:qapp1}

Now $K(\cdot)$ is a symmetric density function with unit variance, and it is easily shown that the term inside the inverse in (25) converges to $\sum_j f_j(x_0)$ times the identity matrix, we have that

$$\widehat{m}_C(x_0) \approx n^{-1} \sum_{i=1}^n \sum_{j=1}^J Z_{ij} K_h(X_{ij}-x_0) \left\{ \sum_{j=1}^J f_j(x_0) \right\}^{-1}. \qquad (26)$$

{eq:qapp2}

Let $\eta_{ij} = Y_{ij} - m(X_{ij})$, $\underline{\eta}_i = (\eta_{i1}, ..., \eta_{iJ})^t$, and define the $nJ \times 1$ vectors

$$\begin{aligned}
\underline{\mathbf{T}}_h(x) &= \left[\{K_h(X_{ij}-x)\}_{j=1}^J \right]_{i=1}^n \left\{ \sum_{j=1}^J f_j(x) \right\}^{-1}; \\
\underline{\eta} &= \left[\{\eta_{ij}\}_{j=1}^J \right]_{i=1}^n.
\end{aligned}$$

¿From the the proof of Theorem 1 (cf. Appendix A.1), we have the expansion

$$\widehat{m}_C(x_0, h) \approx m(x_0) + (1/2)h^2 m^{(2)}(x_0) + (nJ)^{-1} \left\{ \sum_{j=1}^J f_j(x) \right\}^{-1} \sum_{i=1}^n \sum_{j=1}^J \{Y_{ij} - m(X_{ij})\} K_h(X_{ij}-x_0).$$

It then follows that

$$\widehat{m}_C(x_0,h) \approx m(x_0) + (1/2)h^2 m^{(2)}(x_0) + (nJ)^{-1}\underline{\mathbf{T}}_h^t(x_0)\underline{\eta},$$

and hence that

$$\begin{aligned}
\underline{\mathbf{Z}}_i &= \underline{\mathbf{L}}\underline{\mathbf{Y}}_i - (\underline{\mathbf{L}} - \underline{\mathbf{I}})\widehat{\underline{\mathbf{m}}}_{1,pool}(\underline{\mathbf{X}}_i) \\
&= \underline{\mathbf{L}}\underline{\eta}_i + \underline{\mathbf{L}}\underline{\mathbf{m}}(\underline{\mathbf{X}}_i) - (\underline{\mathbf{L}} - \underline{\mathbf{I}})\widehat{\underline{\mathbf{m}}}_W(\underline{\mathbf{X}}_i) \\
&= \underline{\mathbf{L}}\underline{\eta}_i - (nJ)^{-1}(\underline{\mathbf{L}} - \underline{\mathbf{I}}) \begin{bmatrix} \underline{\mathbf{T}}_h^t(X_{i1})\underline{\eta} \\ \cdot \\ \cdot \\ \cdot \\ \underline{\mathbf{T}}_h^t(X_{iJ})\underline{\eta} \end{bmatrix} + \underline{\mathbf{m}}(\underline{\mathbf{X}}_i) - (1/2)h^2(\underline{\mathbf{L}} - \underline{\mathbf{I}})\underline{\mathbf{m}}^{(2)}(\underline{\mathbf{X}}_i). \qquad (27)
\end{aligned}$$

{eq:qapp3}

It is easily seen that the first two terms in (27) contribute only to the variance, while the last two terms contribute only to the bias.

If we write $\underline{\mathbf{L}} = (\ell_{jk})$, then the common diagonal elements are $\ell_{jj} = \ell_d$ and the common off-diagonal elements are $\ell_{jk} = \ell_o$. Then we find that the components of the last two terms in (27) are

$$Z_{ij*} = m(X_{ij}) - (1/2)h^2(\ell_d - 1)m^{(2)}(X_{ij}) - (1/2)h^2\ell_o \sum_{k \neq j} m^{(2)}(X_{ik}).$$

Apply the Z_{ij*} to (26) and take expectations to prove (8).

We now turn to (9). Split the first two terms in (27) into two parts, say Z_{ij1*} and Z_{ij2*}. We first note that (26) when applied to Z_{ij1*} algebraically equals

$$\left\{ n \sum_{j=1}^{J} f_j(x_0) \right\}^{-1} \sum_{i=1}^{n} \{K_h(X_{i1} - x_0), ..., K_h(X_{iJ} - x_0)\} \underline{\mathbf{L}}\underline{\eta}_i. \qquad (28)$$

{eq:qapp4}

Since $\underline{\mathbf{L}}\mathrm{cov}(\underline{\eta}_i)\underline{\mathbf{L}}^t = \tau^2\underline{\mathbf{I}}$, we easily find that (28) has mean zero and approximate variance

$$\tau\gamma \left\{ nh \sum_{j=1}^{J} f_j(x_0) \right\}^{-1}$$

as claimed in (9).

To complete the argument we must show that (26) when applied to Z_{ij2*} is of order $o_p\{(nh)^{-1/2}\}$. The individual terms are

$$Z_{ij2*} = -(nJ)^{-1}(\ell_d - 1)\underline{\mathbf{T}}_h(X_{ij})\underline{\eta} - (nJ)^{-1}\ell_o \sum_{k \neq j} \underline{\mathbf{T}}_h^t(X_{ik})\underline{\eta}.$$

In fact, terms such as $(nJ)^{-1}\underline{\mathbf{T}}_h^t(x)\underline{\eta}$ are very nearly kernel regressions of the η's on the X's evaluated at x, and in (26) these "nearly zero" functions are then averaged via a second kernel operation. Such "double smoothing" has been investigated in other contexts, e.g., Carroll & Wand (1991), and by direct calculation one can show that indeed (26) when applied to Z_{ij2*} is of order $o_p\{(nh)^{-1/2}\}$. This completes the proof.